

AN OVERVIEW OF INVERSE PROBLEM REGULARIZATION USING SPARSITY

J.-L. Starck^a and M. J. Fadili^b,

^a Laboratoire AIM (UMR 7158), CEA/DSM-CNRS-Université Paris Diderot, IRFU, SEDI-SAP, Service d'Astrophysique, Centre de Saclay, F-91191 Gif-Sur-Yvette cedex, France
^b GREYC CNRS-ENSICAEN-Université de Caen, 14050 Caen France.

ABSTRACT

Sparsity constraints are now very popular to regularized inverse problems. We review several approaches which have been proposed in the last ten years to solve inverse problems such as inpainting, deconvolution or blind source separation. We will focus especially on iterative thresholding methods.

Index Terms— Sparsity, Deconvolution, Blind Source Separation, inpainting, Compressed Sensing, iterative thresholding.

1. INTRODUCTION TO SPARSITY

The new sampling theory, **compressed sensing** (also called compressive sensing or compressive sampling), provides an alternative to the well-known Shannon sampling theory [1, 2, 3]. Compressed sensing uses the prior knowledge that signals are sparse, while Shannon theory was designed for frequency band-limited signals. By establishing a direct link between sampling and sparsity, compressed sensing had a huge impact in many scientific fields. A further aspect which has contributed to the success of compressed sensing is that some traditional inverse problems like tomographic image reconstruction can be understood as a compressed sensing problem [3, 4]. Such ill-posed problems need to be regularized, and many different approaches have been proposed in the last 30 years (Tikhonov regularization, Markov random fields, partial differential equations, total variation, wavelets, and so on). But compressed sensing gives a strong theoretical support for methods which seek a sparse solution, since such a solution may be (under appropriate conditions) the exact one. Similar results are hardly accessible with other regularization methods.

By emphasizing so rigorously the importance of sparsity, compressed sensing has also shed light on all work related to sparse data representation (such as the wavelet transform, curvelet transform, etc.). Indeed, a signal is generally not sparse in direct space (i.e. pixel space), but it can be very sparse after being decomposed on a specific set of functions.

1.1. What is Sparsity?

A signal x , $x = [x_1, \dots, x_N]$, is sparse if most of its entries are equal to zero. For instance, a k -sparse signal is a signal where only k samples have a non-zero value. A less strict definition is to consider a signal as weakly sparse or compressible when only a few of its entries have a large magnitude, while most of them are close to zero.

If a signal is not sparse, it may be *sparsified* using a given data representation. For instance, if x is a sine, it is clearly not sparse but its Fourier transform is extremely sparse (i.e. 1-sparse). Hence we say that a signal x is sparse in the Fourier domain if its Fourier coefficients $\hat{x}[u]$, $\hat{x}[u] = \frac{1}{N} \sum_{k=-\infty}^{+\infty} x[k]e^{2i\pi \frac{uk}{N}}$, is sparse. More generally, we can model a vector signal $x \in \mathbb{R}^N$ as the linear combination of T elementary waveforms, also called *signal atoms*: $x = \Phi\alpha = \sum_{i=1}^T \alpha[i]\phi_i$, where $\alpha[i] = \langle x, \phi_i \rangle$ are called the decomposition coefficients of x in the dictionary $\Phi = [\phi_1, \dots, \phi_T]^T$ (the $N \times T$ matrix whose columns are the atoms normalized to a unit ℓ_2 -norm, i.e. $\forall i \in [1, T], \|\phi_i\|_{\ell_2} = 1$).

Therefore to get a sparse representation of our data we need first to define the dictionary Φ and then to compute the coefficients α . x is sparse in Φ if the sorted coefficients in decreasing magnitude have a fast decay; i.e. most of coefficients α vanish but a few.

1.2. What is the Best Dictionary?

Obviously, the best dictionary is the one which leads to the sparsest representation. Hence we could imagine having a huge overcomplete dictionary (i.e. $T \gg N$), but we would be faced with prohibitive computation time cost for calculating the α coefficients. Therefore there is a trade-off between the complexity of our analysis step (i.e. the size of the dictionary) and the computation time. Some specific dictionaries have the advantage of having fast operators and are very good candidates for analyzing the data. The Fourier one is certainly the most famous, but many others have been proposed in the literature such as wavelets [5], ridgelets [6], curvelets [7, 8], bandlets [9], to name only a few.

Thanks to the European Research Council for financial support under grant ERC-228261.

Toward Morphological Diversity

The morphological diversity concept was introduced in [10, 11] in order to model a signal as a finite linear mixture, each component of the mixture being sparse in a given dictionary. The idea is that a single transformation may not always represent an image well, especially if the image contains structures with different spatial morphologies. For instance, if an image is composed of edges and a locally oscillating texture, we can consider edges to be sparse in the curvelet domain while the oscillating texture is better sparsified in the local DCT domain. It has been shown by several authors that choosing a dictionary as a combination of several sub-dictionaries, each sub-dictionary having a fast transformation/reconstruction, allows us to enjoy the advantages of all sub-dictionaries, still having fast and efficient algorithms.

Adaptive representations

Different approaches have also been recently proposed in order to build a dictionary directly from the data. This is the case in learned dictionaries [12], for instance using e.g. the KSVD algorithm [13], the grouplet decomposition [14] or the GMCA method for multichannel/hyperspectral data [15].

2. INVERSE PROBLEMS AND SPARSITY

2.1. The Sparsity Prior

Many image processing problems can be formalized as a linear inverse problem,

$$Y = AX + \varepsilon, \quad (1)$$

, where Y are a set of noisy measurements, ε is an additive noise, X is the solution of our problem, and A is a linear operator. Finding X knowing the data Y and A is an inverse problem. When it has not a unique and stable solution, it is an *ill-posed problem*, and a regularization is necessary to reduce the space of candidate solutions. Once the dictionary Φ is chosen, inverse problems can be regularized used a sparsity penalty. Between all possible solutions, we want the one which has the sparsest representation in the dictionary Φ . Noting α the representation coefficients in Φ , the solution X can be reconstructed as $X = \Phi\alpha$, the sparsity can be measured through the $\|\alpha\|_{\ell^0}$ norm, which indicates the limit of ℓ^p when $p \rightarrow 0$. This counts in fact the number of non-zero elements in the sequence. This approach leads to the following minimization problem :

$$\min_{\alpha} \|\alpha\|_{\ell^0} \text{ s.t. } \|Y - A\Phi\alpha\|_{\ell^2} \leq \sigma. \quad (2)$$

It was proposed to convexify the constraint by substituting the convex ℓ_1 norm for the ℓ_0 norm leading to [16] :

$$\min_{\alpha} \|\alpha\|_{\ell^1} \text{ s.t. } \|Y - A\Phi\alpha\|_{\ell^2} \leq \sigma. \quad (3)$$

This equation can also be recast in its Lagrangian form:

$$\min_{\alpha} \lambda \|\alpha\|_{\ell^1} + \frac{1}{2} \|Y - A\Phi\alpha\|_{\ell^2}^2. \quad (4)$$

Depending on the A operator, there are several ways to obtain the solution of this equation.

2.2. Denoising

The denoising problem corresponds to the case where A is the identity (i.e. $Y = X + N$). Many sparse denoising methods have been proposed in the last ten years based on i) a decomposition of the data on a given dictionary, ii) a thresholding of the coefficients, and iii) a reconstruction of the denoised data. There are also several ways to threshold the coefficients. *Hard thresholding* consists of setting to 0 all coefficients which have an absolute value lower than a threshold λ . Soft thresholding kills or shrinks toward zero the significant coefficients. The most efficient sparse denoising methods do not consider each coefficients independently of his neighborhood, but rather take into account the values of the neighboring coefficients. When the chosen dictionary is the one associated to an orthogonal transform, it is interesting to note that the hard and soft thresholded estimators are the closed-form solutions to the following minimization problems:

$$\tilde{\alpha} = \arg \min_{\alpha} \frac{1}{2} \|Y - \Phi\alpha\|_{\ell^2}^2 + \lambda \|\alpha\|_{\ell^0}^2 \quad \text{hard threshold},$$

$$\tilde{\alpha} = \arg \min_{\alpha} \frac{1}{2} \|Y - \Phi\alpha\|_{\ell^2}^2 + \lambda \|\alpha\|_{\ell^1}^2 \quad \text{soft threshold}.$$

Therefore, hard and soft thresholding have been used as building blocks to derive fast and efficient iterative thresholding techniques to minimize more complex inverse problems with sparsity constraints, such as inpainting and deconvolution.

2.3. Inpainting

The classical image inpainting problem can be defined as follows. Let X be the ideal complete image, Y the observed incomplete image and M the binary mask (i.e. $M_i = 1$ if we have information at pixel i , $M_i = 0$ otherwise). In short, we have: $Y = MX$. Inpainting consists in recovering X knowing Y and M . We thus want to minimize:

$$\min_X \|\Phi^T X\|_0 \quad \text{subject to } Y = MX. \quad (5)$$

Note that we now switch to an analysis-type prior in (5). The solution of such an optimization task can be obtained through an iterative thresholding algorithm called MCA [17, 18] :

$$X^{(n+1)} = \Delta_{\Phi, \lambda_n}(X^{(n)} + Y - MX^{(n)}). \quad (6)$$

where the nonlinear operator $\Delta_{\Phi, \lambda}(Z)$ consists in i) decomposing the signal Z in the dictionary Φ to derive the coefficients $\alpha = \Phi^T Z$, ii) thresholding the coefficients: $\tilde{\alpha} =$

$\rho(\alpha, \lambda)$, where the thresholding operator ρ can either be a hard thresholding or a soft thresholding, and iii) reconstructing \tilde{Z} from the thresholded coefficients $\tilde{\alpha}$.

The threshold parameter λ_n decreases with the iteration number and it plays a role similar to the cooling parameter of the simulated annealing techniques, i.e. it allows the solution to escape from local minima. More details on optimization in inpainting with sparsity can be found in [18]. The case where the dictionary is a union of subdictionaries $\Phi = \{\Phi_1, \dots, \Phi_K\}$ where each Φ_i has a fast operator has also been investigated in [17, 18].

2.4. Deconvolution

In a deconvolution problem, when the sensor is linear, A is the block Toeplitz matrix. A first iterative thresholding deconvolution method was proposed in [19] which consists in the following iterative scheme:

$$X^{(n+1)} = X^{(n)} + A^T \left(\mathbf{WDen}_{\Omega^{(n)}} \left(Y - AX^{(n)} \right) \right) \quad (7)$$

where \mathbf{WDen} is an operator which performs a wavelet thresholding, i.e. applies the wavelet transform of the residual $R^{(n)}$ (i.e. $R^{(n)} = Y - AX^{(n)}$), threshold some wavelet coefficients, and applies the inverse wavelet transform. Only coefficients that belong to the so called *multiresolution support* $\Omega^{(n)}$ [19] are kept, while the others are set to zero. At each iteration, the multiresolution support $\Omega^{(n)}$ is updated by selecting new coefficients in wavelet transform of the residual which have an absolute value larger than a given threshold. The threshold is automatically derived assuming a given noise distribution such as Gaussian or Poisson noise.

More recently, it was shown [20, 21, 22] that the solution of (4) amounts to a thresholded Landweber iteration

$$X^{(n+1)} = \mathbf{WDen}_{\lambda} \left(X^{(n)} + A^T \left(Y - AX^{(n)} \right) \right), \quad (8)$$

with $\|A\| = 1$. In the framework of monotone operator splitting theory, it was shown that for frame dictionaries, a slight modification of this algorithm converges to the solution [22]. Extension to constrained non-linear deconvolution is proposed in [23].

Constraints in the object or image domains

Let us define the *object domain* \mathcal{O} as the space to which the solution belongs, and the *image domain* \mathcal{I} as the space to which the observed data belongs (i.e. if $X \in \mathcal{O}$ then $AX \in \mathcal{I}$). The constraint in (7) was applied in the image domain, while in (8) we have considered constraints on the solution. Hence, two different wavelet based strategies can be chosen in order to regularize the deconvolution problem. The constraint in the image domain through the multiresolution support leads to a very robust way to control the noise. Indeed, whatever the nature of the noise, we can always derive robust

detection levels in the wavelet space and determine scales and positions of the important coefficients. A drawback of the image constraints is that there is no guarantee that the solution is free of artifacts such as ringing around point sources. A second drawback is that image constraints can be used only if the point spread function is relatively compact, i.e. does not smear the information over the whole image.

The property of introducing a robust noise modeling is lost when applying the constraint in the object domain. For example, in the case of Poisson noise, there is no way (except using time consuming Monte Carlo techniques) to estimate the level of the noise in the solution and to adjust properly the thresholds. The second problem with this approach is that, in fact, we try to solve two problems simultaneously (noise amplification and artifact control in the solution) with one parameter (i.e. λ). The choice of this parameter is crucial, while such a parameter is implicit when using the multiresolution support.

Ideally, constraints should be added in both the object and image domains in order to better control the noise by using the multiresolution support and avoid artifact such ringing.

2.5. Sparse Blind Source Separation

In the blind source separation (BSS) setting, the instantaneous linear mixture model assumes that we are given m observations $\{y_1, \dots, y_m\}$ where each $\{y_i\}_{i=1, \dots, m}$ is a row-vector of size t ; each measurement is the linear mixture of n source processes. As the measurements are m different mixtures, source separation techniques aim at recovering the original sources $\mathbf{S} = [s_1^T, \dots, s_n^T]^T$ by taking advantage of some information contained in the way the signals are mixed in the observed data. The linear mixture model is rewritten in matrix form, $\mathbf{Y} = \mathbf{AS} + \mathbf{N}$, where \mathbf{Y} is the $m \times t$ measurement matrix (i.e. observed data), \mathbf{S} is the $n \times t$ source matrix and \mathbf{A} is the $m \times n$ mixing matrix. \mathbf{A} defines the contribution of each source to each measurement. An $m \times t$ matrix \mathbf{N} is added to account for instrumental noise or model imperfections.

It has been shown that sparsity is a very robust regularization to solve BSS for both underdetermined (i.e. we have less observations than unknown, $m < n$) [24] and overdetermined mixtures ($m \geq n$) [25, 15]. The GMCA algorithm [25, 26] finds the sparsest solution through the following iterative scheme:

$$\begin{aligned} S^{(k+1)} &= \Delta_{\Phi, \lambda_k} (A^{+(k)} Y), \\ A^{(k+1)} &= Y S^{(k+1)T} (S^{(k+1)} S^{(k+1)T})^{-1}, \end{aligned} \quad (9)$$

where $A^{+(k)}$ is the pseudo-inverse of the estimated mixing matrix $A^{(k)}$ at iteration k , λ_k is a decreasing threshold and Δ_{Φ, λ_k} is the nonlinear operator which consists in decomposing each source s_i on the dictionary Φ , threshold its coefficients and reconstruct it. Finally, for hyperspectral data, it was advocated to impose sparsity constraints on the columns of the mixing matrix to enhance source recovery [15].

3. REFERENCES

- [1] E.Candès and T.Tao, "Near optimal signal recovery from random projections: Universal encoding strategies?," *IEEE Trans. on Information Theory*, vol. 52, no. 12, pp. 5406–5425, 2006.
- [2] D. Donoho, "Compressed sensing," *IEEE Trans. on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [3] Emmanuel Candès, Justin Romberg, and Terence Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. on Information Theory*, vol. 52, no. 2, pp. 489–509, 2006.
- [4] M. Lustig, D.L Donoho, and J.M Pauly, "Sparse MRI: The application of compressed sensing for rapid mr imaging," *Magnetic Resonance in Medicine*, vol. 58, no. 6, pp. 1182–1195, 2007.
- [5] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, 1998.
- [6] E.J. Candès and D. Donoho, "Ridgelets: the key to high dimensional intermittency?," *Philosophical Transactions of the Royal Society of London A*, vol. 357, pp. 2495–2509, 1999.
- [7] D. Donoho E. Candes, L. Demanet and L.Ying, "Fast discrete curvelet transforms," *SIAM Multiscale Model. Simul.*, 2006, to appear.
- [8] J.-L. Starck, E. Candès, and D.L. Donoho, "The curvelet transform for image denoising," *IEEE Transactions on Image Processing*, vol. 11, no. 6, pp. 131–141, 2002.
- [9] E. Le Pennec and S. Mallat, "Sparse geometric image representations with bandelets," *IEEE Trans. on Image Processing*, vol. 14, no. 4, pp. 423–438, 2005.
- [10] J.-L. Starck, M. Elad, and D.L. Donoho, "Redundant multiscale transforms and their application for morphological component analysis," *Advances in Imaging and Electron Physics*, vol. 132, 2004.
- [11] J.-L. Starck, M. Elad, and D.L. Donoho, "Image decomposition via the combination of sparse representations and a variational approach," *IEEE Transactions on Image Processing*, vol. 14, no. 10, pp. 1570–1582, 2005.
- [12] B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive-field properties by learning a sparse code for natural images," *Nature*, vol. 381, no. 6583, pp. 607–609, June 1996.
- [13] M. Aharon, M. Elad, and A. Bruckstein, "k-svd: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4311–4322, 2006.
- [14] S. Mallat, "Geometrical grouplets," *Applied and Computational Harmonic Analysis (submitted)*, 2007.
- [15] J. Bobin, J.-L. Starck, Y. Moudden, and J. Fadili, "Blind source separation: the sparsity revolution," *Advances in Imaging and Electron Physics*, vol. 152, pp. 221–302, 2008.
- [16] S.S. Chen, D.L. Donoho, and M.A. Saunder, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, pp. 33–61, 1998.
- [17] M. Elad, J.-L. Starck, D. Donoho, and P. Querre, "Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA)," *Journal on Applied and Computational Harmonic Analysis*, vol. 19, pp. 340–358, 2006.
- [18] J. Fadili, J.-L. Starck, and F. Murtagh, "Inpainting and zooming using sparse representations," *Computer Journal*, 2007, in press.
- [19] J.-L. Starck, A. Bijaoui, and F. Murtagh, "Multiresolution support applied to image filtering and deconvolution," *CVGIP: Graphical Models and Image Processing*, vol. 57, pp. 420–431, 1995.
- [20] M.A.T. Figueiredo and R.D. Nowak, "An EM algorithm for wavelet-based image restoration," *IEEE Transactions on Image Processing*, vol. 12, no. 8, pp. 906–916, 2003.
- [21] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Comm. Pure Appl. Math.*, vol. 57, pp. 1413–1541, 2004.
- [22] P. L. Combettes and V. R. Wajs, "Signal recovery by proximal forward-backward splitting," *SIAM Journal on Multiscale Modeling and Simulation*, vol. 4, no. 4, pp. 1168–1200, 2005.
- [23] F.-X. Dupé, M.J. Fadili, and J.-L. Starck, "A proximal iteration for deconvolving poisson noisy images using sparse representations," *Journal of Mathematical Imaging and Vision*, 2008, in press.
- [24] R. Gribonval, H. Rauhut, K. Schnass, and P. Vandergheynst, "Atoms of all channels, unite! average case analysis of multichannel sparse recovery using greedy algorithms," *J. Fourier Analysis and Applications*, vol. 14, pp. 655–687, 2008.
- [25] J. Bobin, J.-L. Starck, J. Fadili, and Y. Moudden, "Sparsity and morphological diversity in blind source separation," *IEEE Transactions on Image Processing*, vol. 13, no. 7, pp. 409–412, 2007.
- [26] J. Bobin, Y. Moudden, M. J. Fadili, and J.-L. Starck, "Morphological diversity and sparsity for multichannel data restoration," *Journal of Mathematical Imaging and Vision*, 2009, in press.